## Supplemental Methods

### Quantitative PCR

Q-PCR was carried out on non-Genomiphi amplified DNA isolated from BAL and OW. TTV primers NG779 (ACWKMCGAATGGCTGAGTTT) and NG781 (CCCKWGCCCGARTTGCCCCT) targeting the UTR of anelloviruses were used (1), and the assay adapted to a quantitative SYBR green-based Q-PCR. Briefly, a 126 bp fragment from TTV was amplified by end-point PCR, TA-cloned into the pCR4 TOPO vector (Invitrogen), quantified by PicoGreen, and used to generate a standard curve. 10µl SYBR green FAST master mix (Applied Biosystems), 50µM of each primer, and 10µl standard curve or DNA (diluted 1:10) were added per well. PCR cycle parameters were as follows: 95°C for 90 seconds, and forty cycles of 94°C for 15 seconds, and 68°C for 1 minute. Melt curve analysis of amplification products showed a single transition suggestive of formation of a single DNA product. All samples were run in duplicate or triplicate and values averaged. The detection limit was 1.4 copies per reaction.

### Bioinformatics Pipeline

Paired-end reads from the MiSeq instrument were quality-trimmed and processed through BMTagger to remove human sequences. Non-human reads were then analyzed using BLAST against the NCBI viral database (downloaded April 2, 2013 from https://www.ncbi.nlm.nih.gov/genomes/Genomes Home.cgi) and the percent of each virus genome covered by reads was calculated. The percentage of paired reads matching the same virus was also calculated. Subsequently, reads were aligned to selected individual viruses one at a time. The alignments were performed using Bowtie2 (2) and visualized using IGV (3). A large number of reads (170,651 across all samples) showed a best BLAST hit from the NCBI viral database to a 50 base region within the RNA plant virus *Physalis mottle virus*. This

sequence covered less than 1% of the genome, and also shared 100% nucleotide identity with DNA cloning vectors, so those sequences were removed manually from the analysis.

Reads were assembled into contigs by iterative deBruijn graph assembly using IDBA-UD (4). Sanger sequencing was performed on selected contigs following PCR amplification using contig-specific primers to verify sequences. Dot plots were generated using Gepard with a word size of 10 nucleotides (5). Anellovirus ORF1 amino acid sequences were identified from our contigs and aligned along with 49 reference TTV, TTMDV, and TTMV sequences from Genbank. Sequences were aligned using MUSCLE v.3.8.31 (6) and trimmed to include only columns with 80% amino acid coverage. Subsequently, the 10% most variable columns were removed. The phylogenetic tree was built using FastTree v. 2.1.3 (7) which calculates local support values using the Shimodaira-Hasegawa (SH) test to estimate the reliability of each split compared to alternate topologies (7-10). The groups, color bars corresponding to the samples, and SH-indices were incorporated using iTOL (11). Groups were chosen based on clades determined by SH-indices (>90%) and a guide tree previously published by Okamoto et al (12).

References:


1. Ninomiya M, Takahashi M, Nishizawa T, Shimosegawa T, Okamoto H. Development of pcr assays with nested primers specific for differential detection of three human anelloviruses and early acquisition of dual or triple infection during infancy. *Journal of clinical microbiology* 2008;46:507-514.
2. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* 2009;10:R25.
3. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nature biotechnology* 2011;29:24-26.
4. Peng Y, Leung HC, Yiu SM, Chin FY. Idba-ud: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012;28:1420-1428.
5. Krumsiek J, Arnold R, Rattei T. Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 2007;23:1026-1028.
6. Edgar RC. Muscle: Multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 2004;32:1792-1797.
7. Price MN, Dehal PS, Arkin AP. Fasttree 2--approximately maximum-likelihood trees for large alignments. *PloS one* 2010;5:e9490.

8. Shimodaira HH, M. Multiple comparisons of log-liklihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 1999;16:1114-1116.
9. Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with phyml. *Methods in molecular biology* 2009;537:113-137.
10. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of phyml 3.0. *Systematic biology* 2010;59:307-321.
11. Letunic I, Bork P. Interactive tree of life (itol): An online tool for phylogenetic tree display and annotation. *Bioinformatics* 2007;23:127-128.
12. De Villiers E-M, Zur Hausen H. Tt viruses: The still elusive human pathogens. Springer; 2009.